

Aus dem Institut für Klinische Chemie und Laboratoriumsmedizin
(Direktor Univ.-Prof. Dr. med. Matthias Nauck)
der Universitätsmedizin der Ernst-Moritz-Arndt-Universität Greifswald

Häufigkeit und Ausmaß von Messabweichungen
ermittelt mit Hilfe von Doppelmessungen

Inaugural-Dissertation
zur
Erlangung des akademischen
Grades
Doktor der Medizin
(Dr.med.)
der
Universitätsmedizin
der
Ernst-Moritz-Arndt-Universität
Greifswald
2017

vorgelegt von:
Stefanie Mareike Neubig
geb. am: 27.02.1986
in: Heidelberg

Dekan: Univ.-Prof. Dr. rer. nat. Max P. Baur

1. Gutachter: Univ.- Prof. Dr. med. Matthias Nauck

2. Gutachter: PD Dr. med. Thomas Streichert

Ort, Raum: Laboratoriumsmedizin, Universitätsmedizin Greifswald,
Besprechungsraum 10.0.16

Tag der Disputation: 20.04.2018

Für meine Eltern

INHALTSVERZEICHNIS

	SEITE
1. EINLEITUNG	1
1.1. Qualitätssicherung laboratoriumsmedizinischer Untersuchungen	1
1.2. Impräzision und Unrichtigkeit	2
1.3. Doppelmessungen als innovatives Element zur Beschreibung analytischer Qualität	2
1.4. Zielsetzung	6
2. MATERIAL UND METHODEN	7
2.1. Proben	7
2.2. Probenfluss und Laboranalysen	7
2.3. Qualitätskontrollen	8
2.4. Statistik	8
3. ERGEBNISSE	10
4. DISKUSSION	15
5. ZUSAMMENFASSUNG	18
6. LITERATURVERZEICHNISS	19
7. ANHANG	22
A1 Abbildungsverzeichnis	22
A2 Tabellenverzeichnis	23
A3 Publikation	24

Abkürzungsverzeichnis

ALAT	Alanin-Aminotransferase
ASAT	Aspartat-Aminotransferase
AZ95	A-Zone 95 %
A-Zone	Zone of acceptance
Ca	Calcium
CAP	College of American Pathologists
Chol	Cholesterol
CLIA	Clinical Laboratory Improvement Amendments
CLSI	Clinical and Laboratory Standards Institute
CRP	C-reaktives Protein
Krea	Kreatinin
Lak	Laktat
MD	Minimal Difference
MPBetreibV	Medizinprodukte-Betreiberverordnung
OPM	Outlier per mille
QK	Qualitätskontrolle
Rili-BÄK	Richtlinie der Bundesärztekammer zur Qualitätssicherung laboratoriumsmedizinischer Untersuchungen
SD	Standardabweichung
Trig	Triglyceride
TSH	Thyreidea-stimulierendes Hormon
VK	Variationskoeffizient

1. Einleitung

1.1. Qualitätssicherung laboratoriumsmedizinischer Untersuchungen

Die regelmäßige Durchführung von Qualitätskontrollen stellt ein zentrales Element der Qualitätssicherung für laboratoriumsmedizinische Untersuchungen dar (1). Etablierte Vorgaben zu Frequenz, Durchführung und Bewertung von internen Qualitätskontrollen finden sich z.B. bei Westgard (2). In Deutschland gilt für alle laboratoriumsmedizinischen Untersuchungen in der Heilkunde rechtlich verbindlich die Richtlinie der Bundesärztekammer zur Qualitätssicherung laboratoriumsmedizinischer Untersuchungen (Rili-BÄK) (3). Neben der generellen Verpflichtung der Leistungserbringer der Krankenkassen zur Qualitätssicherung im § 135a des Sozialgesetzbuches (Fünftes Buch) ist die regelmäßige Qualitätssicherung in der Laboratoriumsmedizin im § 9 der Medizinprodukte-Betreiberverordnung (MPBetreibV) festgeschrieben. Dort wird auf die Rili-BÄK verwiesen, die durch die Bundesärztekammer vorgegeben wird. Die erste Fassung der Rili-BÄK wurde 1971 veröffentlicht. Weitere überarbeitete Versionen folgten, die zunehmend auf mehr Anwendungsgebiete ausgeweitet wurden. Inzwischen umfasst die Rili-BÄK neben dem allgemeinen Teil A fünf spezielle Teile B1 bis B5 (3). Im Rahmen dieser Arbeit werden ausschließlich Analyte betrachtet, die in den Teil B1 „Quantitative Untersuchungen“ fallen.

Während bei der externen Qualitätssicherung in der Regel vier Mal jährlich pro Analyt Ringversuche durchgeführt werden, werden im Rahmen der internen Qualitätssicherung mindestens zwei Mal täglich pro Gerät und Analyt Kontrollproben-einzelmessungen mit Qualitätskontrollmaterialien durchgeführt. Die Bewertungskriterien der internen Qualitätssicherung umfassen unter anderem die zulässige relative Abweichung des Einzelwertes beziehungsweise rückblickend auf einen Kontrollzyklus den relativen quadratischen Mittelwert. Beide Angaben sind in der sogenannten „Spalte 3“ der Tabelle B1 festgelegt. Für quantitative Analyte, die nicht in dieser Tabelle aufgeführt sind, gelten für die Bewertung der Qualitätskontrollen zunächst die Herstellerangaben bis eine ausreichende Anzahl an Messwerten für die Ermittlung von laboratoriumsinternen Grenzen zur Verfügung

stehen. Es werden für jeden Assay täglich mindestens zwei Kontrollproben mitgeführt. Überschreitet ein Kontrollprobeneinzelwert die festgelegte Fehlergrenze, so ist das Messverfahren für Untersuchungen von Patientenproben gesperrt und die Ursache der Überschreitung muss eruiert und beseitigt werden, bevor das Messverfahren wieder freigegeben wird.

1.2. Impräzision und Unrichtigkeit

Messfehler können mit Hilfe der Impräzision und Unrichtigkeit beschrieben werden (4). Dabei beschreibt die Impräzision die Streuung der Messwerte in einem Verfahren und wird mit der Standardabweichung (SD) bzw. dem Variationskoeffizienten (VK) erfasst. Für den Einsender kann die Impräzision verständlicher als Minimal Difference (MD) ausgedrückt werden. Die MD beschreibt auf Grundlage der Standardabweichung bei einem gegebenen Konfidenzintervall die geringste Differenz zwischen zwei Messungen, um diese als voneinander verschieden zu identifizieren (5). Die Impräzision kann mit Hilfe der internen Qualitätskontrollen erfasst werden.

Die Unrichtigkeit beschreibt eine systematische Messabweichung (6). In Laboratorien, die gemäß Rili-BÄK arbeiten, kann diese ebenfalls, zumindest bedingt, mit Hilfe der internen Qualitätskontrolle erfasst werden, da die Qualitätskontrollzielwerte einheitlich pro Messsystem für ganz Deutschland gelten. Für Laboratorien, die zum Beispiel gemäß der Westgard-Regeln arbeiten, können für die Erfassung der Unrichtigkeit nur die externen Qualitätskontrollen herangezogen werden, da hier jedes Laboratorium eigene Zielwerte für die internen Qualitätskontrollen ermittelt. Unabhängig davon ist auch in der Rili-BÄK für viele Analyte eine Pflicht zur Durchführung externer Qualitätskontrollen (Ringversuche) vorgeschrieben.

1.3. Doppelmessungen als innovatives Element zur Beschreibung analytischer Qualität

Während früher Doppelmessungen ein weit verbreitetes Instrument waren um die Impräzision eines Messergebnisses zu reduzieren, werden in der modernen, hochautomatisierten Laboratoriumsmedizin fast ausschließlich Einfachmessungen durchgeführt. Gründe sind zum einen die zusätzlichen Kosten, aber auch die dadurch entstehende Zeitverzögerung bis zur Ergebnismitteilung bei einer nur

geringen Anzahl von Messabweichungen, die hierdurch aufgedeckt werden würden (7,8). Auch wenn interne Qualitätskontrollen ein weithin akzeptiertes Konzept zur Qualitätssicherung laboratoriumsmedizinischer Untersuchungen darstellen, so bleibt bezüglich der analytischen Qualität einer einzelnen Patientenprobenmessung eine Restunsicherheit (9–13), die auch abhängt von der Anzahl der internen Qualitätskontrollen im Verhältnis zur Anzahl der Patientenprobenmessungen. Die zwischen den Kontrollproben liegenden Messungen von Patientenproben können je nach Auslastung des Laboratoriums bis zu vielen tausend betragen. Mit einer geringen Frequenz auftretende Messabweichungen bei Patientenproben können somit weder in ihrer Anzahl noch hinsichtlich ihres Ausmaßes erfasst werden (14). In der Literatur sind bereits Fälle beschrieben, in denen durch falsche Laborergebnisse falsche Diagnosen gestellt und dadurch unnötige und für den Patienten belastende Therapien eingeleitet wurden (15–24). Regelhaft durchgeführte Doppelmessungen hingegen können zufällige, das heißt nicht systematisch auftretende Messfehler aufdecken bevor das Ergebnis an den behandelnden Arzt berichtet wird.

Die Definition, welcher Unterschied zwischen zwei Doppelmessungen als Fehler beziehungsweise Messabweichung gewertet wird, ist Gegenstand intensiver Diskussionen (1, 2, 4, 6–12, 14, 25–28). Häufig wird hier ein Vielfaches der Standardabweichung zur Definition herangezogen (9–12, 29). Des Weiteren können – statt einer festen Definition einer Messabweichung – individuelle Definitionen spezifisch für einen Analyten angewendet werden, um die klinischen Anforderungen und das technisch Realisierbare eines jeweiligen Assays zu berücksichtigen. Die uneinheitliche Auffassung was als Messfehler zu werten ist, führt zu sehr heterogenen Berichten von Fehlerhäufigkeiten in der Literatur. Hierdurch wird eine Vergleichbarkeit der Fehlerhäufigkeiten verschiedener Systeme erschwert. Petersmann et al. untersuchten die Fehlerhäufigkeit bei einem Glucose-Assay unter Anwendung fließender, durch den Anwender festlegbare Grenzen mit Hilfe von Doppelmessungen.

In der vorliegenden Studie wird das von Petersmann et al. beschriebene A-Zonen-Modell (zone of acceptance) angewendet, das ähnlich den Error-Grid-Analysen Werte als fehlerhaft oder nicht fehlerhaft klassifiziert (30–33). Durch die Variation der A-Zonen-Weite ermöglicht dieses Modell eine graphische Darstellung der Fehlerhäufigkeiten in Abhängigkeit von der Fehlerdefinition.

Bei der A-Zone werden Wertabweichungen der Doppelmessungen für jeden Assay mittels Regressions- und Korrelationsanalyse in einem Streudiagramm dargestellt, bei dem die X-Achse den zuerst gemessenen Wert und die Y-Achse den als zweites gemessenen Wert angibt. Übereinstimmungen der Doppelmessungen liegen auf der Äquivalenzlinie (durchgezogene Linie, Abb. 1). Messabweichungen zeigen sich außerhalb der Äquivalenzlinie. Eine symmetrische Zone (gestrichelte Linie, Abb. 1) um die Äquivalenzlinie bildet die A-Zone. Eine A-Zonen-Weite von 5 % bedeutet beispielsweise, dass Werte, die sich weniger als 5 % vom ersten Messwert unterscheiden, innerhalb dieser Zone liegen (grüne Dreiecke, Abb. 1), während Abweichungen $> 5\%$ außerhalb der A-Zone liegen und in diesem Falle als Messabweichung gewertet werden (rotes Viereck, Abb. 1).

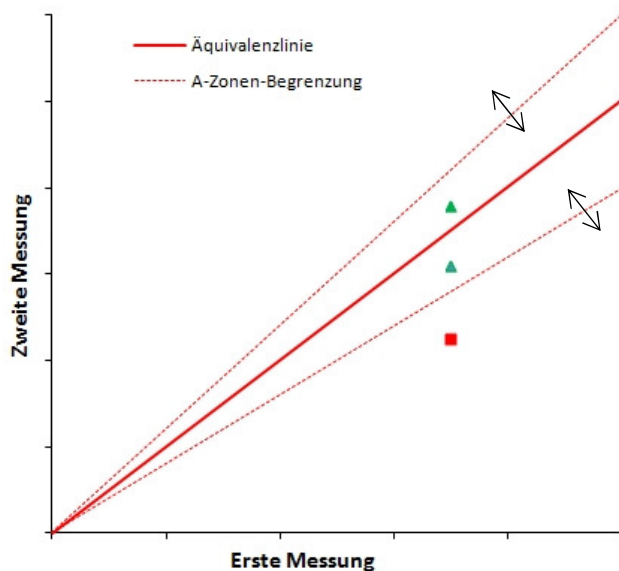


Abbildung 1: A-Zonen-Modell.

Messergebnisse einer Doppelmessung werden in dem Streudiagramm abgebildet. Die A-Zone bildet eine symmetrische Zone um die Äquivalenzlinie, dessen Weite variiert werden kann (schwarzen Pfeile). Übereinstimmende Doppelmessungen zeigen sich auf der Äquivalenzlinie. Messabweichungen, die die durch die A-Zonen-Weite definierte Differenz nicht überschreiten, liegen innerhalb der A-Zonen-Begrenzungslinien (grüne Dreiecke), während größere Messabweichungen außerhalb der A-Zone liegen (rotes Vierecke).

Durch Variation der symmetrischen A-Zonen-Weite um die Äquivalenzlinie kann die erlaubte Fehlergrenze angepasst werden (Pfeile, Abb. 1). Eine geringe A-Zonen-Weite bedeutet eine prozentual kleine zulässige Messabweichung. Dadurch gelten bereits geringfügig unterschiedliche Messergebnisse als Messabweichungen.

Dagegen bedeutet eine große A-Zonen-Weite, dass erst prozentual große Messergebnisunterschiede als Messabweichungen angesehen werden. Petersmann et al. ermittelte mittels A-Zonen-Modell für Glucose eine Frequenz von Messabweichungen von 0,3–25 pro 1000 Messungen bei Weiten der A-zone zwischen 2 und 11 % (14).

Dieses Wissen über die Häufigkeit und das Ausmaß von Messabweichungen kann bei der Bewertung des Messverfahrens für das Labor hilfreich sein, ob es für den angestrebten klinischen Zweck eine angemessene Qualität liefert.

1.4. Zielsetzung

Ziel der vorliegenden Promotionsarbeit ist es, anhand einer großen Anzahl von kontinuierlich durchgeführten Doppelmessungen bei neun häufig genutzten Assays die Häufigkeit und das Ausmaß von Messabweichungen unter Routinebedingungen zu ermitteln. Dabei soll das von Petersmann et al. verwendete Konzept für die Verwendung in medizinischen Laboratorien weiterentwickelt werden. Innovative Qualitätsmarker sollen abgeleitet und mit einer guten Datenbasis unterlegt werden.

2. Material und Methoden

2.1. Proben

Von Dezember 2011 bis Dezember 2012 wurden im Institut für Klinische Chemie und Laboratoriumsmedizin der Universitätsmedizin Greifswald alle eingehenden Patientenproben, für die eine oder mehrere Anforderungen der neun untersuchten Analyten vorlagen, doppelt gemessen. Erfasst wurden die katalytische Aktivitätskonzentration von Alanin-Aminotransferase (ALAT) und Aspartat-Aminotransferase (ASAT) sowie die Konzentration von Calcium, Cholesterol, C-reaktivem Protein (CRP), Kreatinin, Laktat, Triglyceriden und Thyreoidea-stimulierendem Hormon (TSH). Die Proben stammten von einer unselektierten Population von ambulanten und stationären Patienten. Ein positives Votum der Ethikkommission der Ärztekammer Mecklenburg-Vorpommern bei der Ernst–Moritz–Arndt-Universität Greifswald zur Verwendung von Restmaterialien (Registrierungsnummer III UV 39/03) besteht.

2.2. Probenfluss und Laboranalysen

Alle eingehenden Proben wurden mit dem Laborautomationssystem StreamLAB[®] (Siemens Healthcare Diagnostics, Eschborn, Deutschland) zufällig einem von drei Dimension Vista[®] 1500 Analysesystemen (Siemens Healthcare Diagnostics, Eschborn, Deutschland) zugeführt. Hier erfolgte dann automatisch die Durchführung der Doppelmessung. Messprozedur und die hierfür verwendeten Reagenzien richteten sich nach den Herstellervorgaben (Siemens Healthcare Diagnostics, Eschborn, Deutschland). Nur das Ergebnis der ersten Messung wurde für die Patientenversorgung verwendet und an das Laborinformationssystem gesendet. Für das Laborpersonal waren die Ergebnisse der zweiten Messung nicht sichtbar und wurden demzufolge auch nicht für die Beurteilung der Patientenergebnisse berücksichtigt. Die Studiendaten wurden separat anhand der Auftragsnummer ohne weitere Patienten-identifizierenden Daten in regelmäßigen Abständen direkt von den Geräten gewonnen und waren somit zu jedem Zeitpunkt anonymisiert.

Einschlusskriterien der Messergebnisse waren:

- gültige Qualitätskontrolle gemäß Rili-BÄK

- Konzentration im Gültigkeitsbereich der Rili-BÄK (Tabelle B1, Spalte 4) oder der Herstellerangaben, wenn dieser Bereich enger war
- ausreichend Restmaterial für die Doppelmessung.

2.3. Qualitätskontrollen

Die interne Qualitätssicherung erfolgte nach den Vorgaben der Rili-BÄK anhand von mindestens zwei Qualitätskontrollprobeneinzelmessungen mit Kontrollmaterial an jedem Messtag sowie nach jedem Eingriff ins Labormesssystem, wie z.B. Reparatur oder Kalibration. Die für die Kontrollmessungen verwendeten Qualitätskontrollmaterialien waren für ALAT, ASAT, Calcium, Cholesterol, Kreatinin, Laktat und Triglyceride TRU-Liquid Monitrol (Charge 1AQ104) und für TSH Liquimmune (Charge 9LQ105, beides von Thermo Fisher Scientific, Schwerte, Deutschland), sowie für CRP Protein2 (Charge 1LQH01, von Siemens Healthcare Diagnostics, Eschborn, Deutschland).

2.4. Statistik

Die Datenanalyse erfolgte mit Microsoft Excel® (2010). Analog zu dem von Petersmann et al. angewendeten Modell werden die Ergebnisse der Doppelmessungen in einem Streudiagramm mit Regressions- und Korrelations-Analyse dargestellt. Messergebnisse außerhalb der A-Zone wurden als Messabweichung angesehen. Durch Variierung der A-Zonen-Weite konnte die Anzahl der Messabweichungen bei verschiedenen Fehlergrenzen erfasst werden.

In dieser Studie wurde bei allen Analyten die A-Zonen-Weite schrittweise in 1 %-Stufen von 1 % bis auf 14 % erhöht und die Anzahl der Messabweichungen bei den verschiedenen A-Zonen-Weiten erfasst und grafisch dargestellt (Abb. 2). Alle Analyten hatten ab einer A-Zonen-Weite von 12 % weniger als 50 von 1000 Messungen außerhalb der A-Zone. Da die Kurven im weiteren Verlauf einen asymptotisch abfallenden Verlauf zur X-Achse zeigten, wurden die Kurven exemplarisch bis zu einer A-Zonen-Weite von 14 % aufgezeigt, danach allerdings auf eine weitere Darstellung verzichtet. Zum Vergleich der analytischen Leistungsfähigkeit verschiedener Assays wurden zwei neue Qualitätsmarker mit Hilfe den Studiendaten etabliert:

1. AZ95 (A-Zone 95 %): Dieser Wert beschreibt die *A-Zonen-Weite* eines Analyten, bei dem 95 % der Doppelmessungen innerhalb der A-Zone liegen und somit bei dieser A-Zonen-Weite 5 % Messabweichungen bestehen. Hat ein Analyt beispielsweise einen AZ95 von 3 %, bedeutet dies, dass 95 % aller Messungen eine maximale Messabweichung von 3 % zu erwarten ist und 5 % aller Messungen eine größere Messabweichung als 3 % besitzen.

Der AZ95 des jeweiligen Analyten kann in Abb. 2 an den Werten auf der X-Achse an dem Punkt abgelesen werden, wo die rote horizontale Linie die Analytenkurve kreuzt. Diese rote horizontale Linie kennzeichnet die Stelle, an der 50 Messausreißer bei 1000 Messungen (d.h. 5 % Messabweichungen) auftreten.

Diese Werte jedes Analyten wurden durch Least-Squares-Fitting nach einem Polynom 5. Ordnung der Form $y = \sum_{n=0}^5 a_n \cdot x^{-n}$ approximiert.

2. OPM (outlier per mille): Die *Anzahl von Messabweichungen* bei 1000 Messungen bei einer festgelegten A-Zonen-Weite von 5 %. Diese Anzahl kann in Abb. 2 an der Y-Achse an der Kreuzung der Analytenkurve mit der vertikalen roten Linie abgelesen werden.

Analyten, bei denen 95 % der Messungen nicht mehr als 5 % abweichen, verlaufen durch den linken unteren Quadranten der Abbildung (Abb. 2, grau unterlegter Quadrant). Diese Grenzen sind unter Berücksichtigung des CLSI gewählt, das für die meisten Messverfahren erfordert, dass 95 % der Messungen innerhalb der A-Zone liegen sollten (30).

3. Ergebnisse

Insgesamt wurden 237.261 Doppelmessungen bei 9 verschiedenen Analyten untersucht. Pro Analyt wurden zwischen 1.596 (Laktat) und 73.242 (Kreatinin) Doppelmessungen durchgeführt. Bei 2 Analyten (Calcium, Kreatinin) wurden jeweils über 65.000 Doppelmessungen eingeschlossen. Bei 5 Analyten (ASAT, Cholesterol, CRP, Laktat, Triglyceriden, TSH) bewegte sich die untersuchte Anzahl zwischen 11.000 – 27.800. Nur 2 der Analyte hatten mit 5.620 Proben für ALAT und 1.596 Proben für Laktat weniger als 6.000 Doppelmessungen (Tabelle 3).

Tabelle 1: Erhebungszeitraum.

Analyt	Messzeitraum
ALAT	01.02.2012–09.4.2012
ASAT	01.01.2012–09.04.2012
Calcium	19.04.2012–11.12.2012
Cholesterol	19.04.2012–11.12.2012
CRP	18.04.2012–19.12.2012
Kreatinin	18.04.2012–19.11.2012
Laktat	19.04.2012–29.11.2012
Triglyceride	18.04.2012–11.12.2012
TSH	19.04.2012–29.11.2012

Tabelle 2: Regressionsanalyse.

Analyt	Lineare Regression
ALAT	$Y = 0,999 * X + 0,0 \mu\text{katal/L}$
ASAT	$Y = 0,999 * X + 0,0 \mu\text{katal/L}$
Calcium	$Y = 0,972 * X + 0,062 \text{ mmol/L}$
Cholesterol	$Y = 0,990 * X + 0,015 \text{ mmol/L}$
CRP	$Y = 0,997 * X + 0,053 \text{ mg/L}$
Kreatinin	$Y = 0,998 * X + 0,114 \mu\text{mol/L}$
Laktat	$Y = 0,995 * X - 0,002 \text{ mmol/L}$
Triglyceride	$Y = 0,984 * X + 0,024 \text{ mmol/L}$
TSH	$Y = 1,000 * X - 0,001 \text{ mU/L}$

Tabelle 3: Übersicht der untersuchten Analyten.

Für jeden Analyten ist der Messbereich gemäß Herstellerangaben, der Gültigkeitsbereich der Rili-BÄK (Tabelle B1, Spalte 4), die Anzahl der durchgeführten Doppelmessungen sowie die Impräzision, ermittelt anhand der Qualitätskontrollen (QK) in einer niedrigen und einer hohen Konzentration, aufgeführt.

Analyt [Einheit]	Messbereich gemäß Hersteller- angaben	Gültig- keits- bereich Rili-BÄK	Anzahl Doppel- Messungen	VK (QK niedrig)	VK (QK hoch)
ALAT [μkat/L]	0.1 - 16.7	0.5 – 5.0	5620	3.1 - 5.0 % (0.7 μkat/L)	2.8 - 3.6 % (1.75 μkat/L)
ASAT [μkat/L]	0.05 - 16.7	0.33 – 6.67	11797	2.1 - 5.7 % (0.6 μkat/L)	0.8 - 1.5 % (3.23 μkat/L)
Calcium [mmol/L]	1.25 - 3.75	1 – 6	65077	1.1 - 2.4 % (2.41 mmol/L)	1.0 – 2.2 % (2.85 mmol/L)
Cholesterol [mmol/L]	1.29 - 15.54	1.3 – 9.1	12092	1.7 - 5,2% (5.1 mmol/L)	1.8 - 5.4 % (2.74 mmol/L)
Kreatinin [μmol/L]	9 – 1768	44 – 884	73242	2.2 - 7.5 % (118 μmol/L)	0.9 - 3.8 % (629 μmol/L)
CRP [mg/L]	3.1 – 190	1 – 120	27836	2.4 - 6.1 % (52 mg/L)	2.3 - 7.0 % (12.1 mg/L)
Laktat [mmol/L]	0.1 – 15	1 – 10	1596	1.2 - 4.4 % (2.42 mmol/L)	1.2 - 8.3 % (5.71 mmol/L)
Triglyceride [mmol/L]	0.02 - 11.3	0.68 – 4.6	17094	1.3 - 3.8 % (2.35 mmol/L)	1.6 - 4.5 % (1.35 mmol/L)
TSH [mU/L]	0.005 – 100	0.1 - 40	23321	2.0 - 4.3 % (0.27 mU/L)	1.8 - 3.6 % (16.4 mU/L)

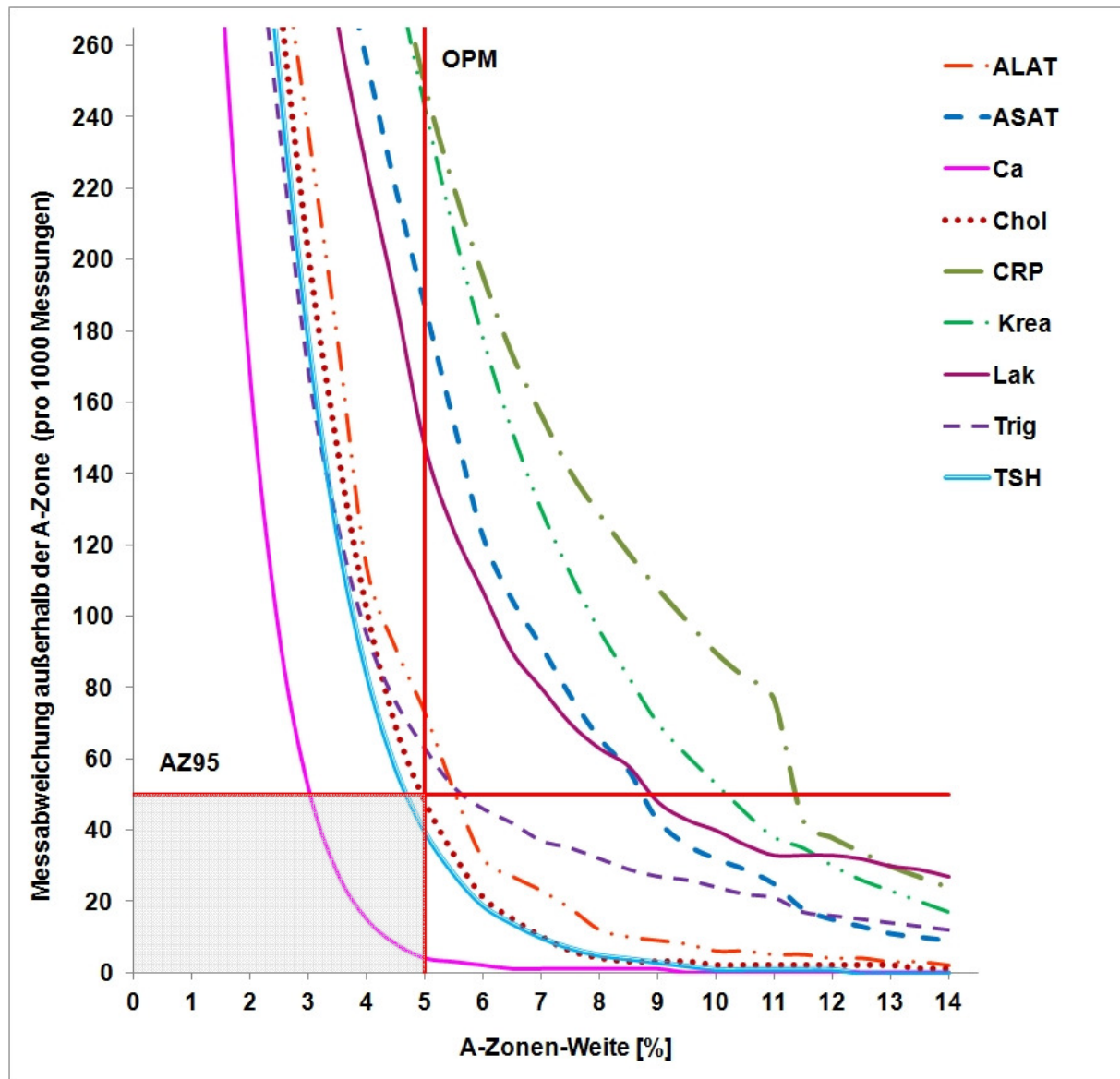


Abbildung 2: Anzahl der Messabweichungen (bei 1000 Doppelmessungen) in Relation zur A-Zonen-Weite.

An den Schnittpunkten der horizontalen Kennlinie mit den Analytenkurven sind die AZ95-Werte auf der X-Achse ablesbar. Der OPM ergibt sich auf der Y-Achse aus den Schnittpunkten der vertikalen Kennlinie mit den Analytenkurven. Der grau hinterlegte Bereich im linken unteren Quadranten stellt den Bereich dar, in dem Analyten sowohl einen AZ95 von maximal 5 % bei gleichzeitig einem OPM von maximal 50 besitzen. Alanin-Aminotransferase (ALAT), Aspartat-Aminotransferase (ASAT), Calcium (Ca), Cholesterol (Chol), C-reaktives Protein (CRP), Kreatinin (Krea), Laktat (Lak), Triglyceride (Trig), Thyreoidea-stimulierendes Hormon (TSH).

Die Werte des AZ95 bewegten sich zwischen 3,2 % (Calcium) und 11,3 % (CRP). Hier zeigte der Calcium-Assay die beste analytische Leistungsfähigkeit mit einem AZ95 von 3,2 %. Insgesamt hatten 5 der 9 Analyte einen AZ95 unter 6 % (Calcium, TSH, Cholesterol, ALAT und Triglyceride). Die übrigen Assays wiesen einen AZ95 zwischen 8,7 % (ALAT) und 11,3 % (CRP) auf (siehe Abb.2 und 3).

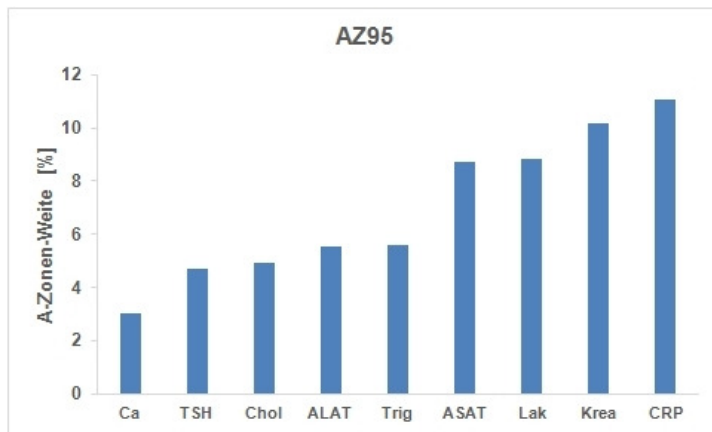


Abbildung 3: Der AZ95 in aufsteigender Reihenfolge.

Die A-Zonen-Weite, bei der 95 % der Doppelmessungen innerhalb der A-Zone lagen für Calcium (Ca), Thyreoidea-stimulierendem Hormon (TSH), Cholesterol (Chol), Alanin-Aminotransferase (ALAT), Triglyceride (Trig), Aspartat-Aminotransferase (ASAT), Laktat (Lak), Kreatinin (Krea), C-reaktivem Protein (CRP).

Wird der OPM als Parameter für die analytische Leistungsfähigkeit angewendet, zeigt erneut Calcium die besten Werte mit einem OPM von 5 (Abb. 4). Calcium, TSH, Cholesterol, Triglyceride und ALAT zeigten mit unter 100 deutlich weniger Messabweichungen pro 1000 Messungen bei einer A-Zonen-Weite von 5 % als ASAT, Laktat, Kreatinin und CRP. Die größte Anzahl von Messausreißern bei einer A-Zonen-Weite von 5 % wird von CRP und Kreatinin erreicht mit einem OPM bis zu 250, d.h. bis zu 250 Messausreißer (siehe Abb. 2 und 4).

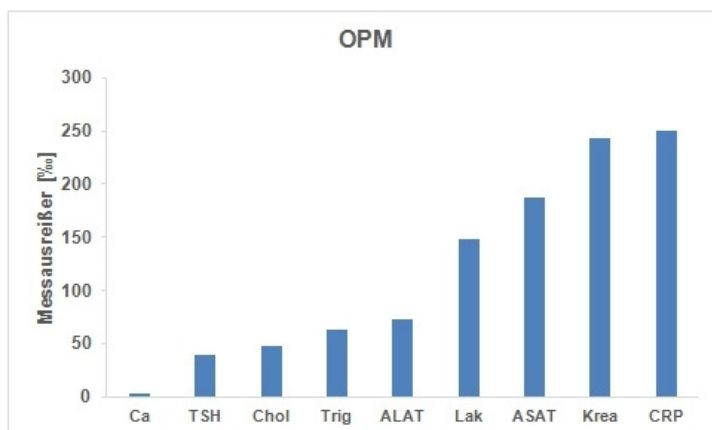


Abbildung 4: Der OPM in aufsteigender Reihenfolge.

Die Anzahl der Messausreißer pro 1000 Messungen bei einer A-Zonen-Weite von 5 % für Calcium (Ca), Thyreoidea-stimulierendem Hormon (TSH), Cholesterol (Chol), Triglyceride (Trig), Alanin-Aminotransferase (ALAT), Laktat (Lak), Aspartat-Aminotransferase (ASAT), Kreatinin (Krea), C-reaktivem Protein (CRP).

Nur 3 der untersuchten Analyten erreichten gleichzeitig einen $AZ95 < 5 \%$ sowie einen $OPM < 50$: Calcium, TSH und Cholesterol. Die Kurven dieser Assays durchlaufen in der Abb. 2 den linken unteren Quadranten (Abb. 2, grau unterlegter Quadrant).

4. Diskussion

In der vorliegenden Studie wurden mit Hilfe einer großen Anzahl von Doppelmessungen unter Routinebedingungen auftretende Messabweichungen bei neun häufig gemessenen Analyten ermittelt und beschrieben. Die so ausgewerteten Messabweichungen geben Informationen über die Leistungsfähigkeit und analytische Robustheit von Messverfahren.

Auf dieser Basis wurden zwei neue Qualitätsmarker für die analytische Leistungsfähigkeit entwickelt (AZ95 und OPM):

Der AZ95 beschreibt die Weite der A-Zone, bei der 95 % der Messergebnisse innerhalb der A-Zone liegen (Schnittpunkt der Analytenkurve mit horizontaler Linie, abgelesen auf der Y-Achse, Abb.2). Dieser Marker wurde in Anlehnung an das 95 %-Konfidenzintervall gewählt. Die Werte des AZ95 lagen bei den untersuchten Analyten zwischen 3,2 % (Calcium) und 11,3 % (CRP). Je kleiner der AZ95 eines Messverfahrens ist, desto geringere Messabweichungen sind bei den Messungen zu erwarten. Die Bedeutung des AZ95 wird anhand unserer Ergebnisse für Calcium und Kreatinin durch Übersetzung der relativen in absolute Zahlen veranschaulicht: Der AZ95 von 3,2 % für Calcium bedeutet, dass bei einer Calcium-Konzentration von 2 mmol/l Wertabweichungen von maximal $\pm 3,2$ % bei 95 % aller Messungen erwartet werden können. Das bedeutet bei einer Doppelmessung, dass bei einer Konzentration von 2 mmol/l der 2. Messwert in 95 % der Messungen zwischen 1,94 und 2,06 mmol/l liegt. Eine hiermit vergleichbare analytische Leistungsfähigkeit wurde bei Glucosemessungen in einer vorherigen Studie mit derselben Verfahrensweise ermittelt, die eine A-Zonen-Weite von 4 % ermittelte, um 95 % der Messungen in die A-Zone einzuschließen (14). Für Kreatinin wurde in der vorliegenden Studie ein AZ95 von 10 % ermittelt. Bei einer Kreatinin-Konzentration von 100 $\mu\text{mol/l}$ liegen bei einer Wiederholungsmessung 95 % der Werte zwischen 90 und 110 $\mu\text{mol/l}$. Bei den übrigen 5 % der Messungen muss mit noch größeren Messabweichungen gerechnet werden.

Der OPM gibt die Anzahl von Messausreißern bei 1000 Doppelmessungen bei einer fixen A-Zonen-Weite von 5 % an (Schnittpunkt der Analytenkurve mit vertikaler Linie, Abb. 2). Bei einem OPM von 250, wie dies für CRP ermittelt wurde, bedeutet dies bei einer Zielkonzentration von 5 mg/l folglich 250 Messabweichungen von mehr als \pm

0,25 mg/l bei 1000 Messungen. Für Kreatinin, dessen OPM bei 243 gefunden wurde, entspricht dies 243 Messabweichungen, die bei einer Konzentration von 100 $\mu\text{mol/l}$ mehr als $\pm 5,0 \mu\text{mol/l}$ betragen. Calcium wies von allen untersuchten Analyten den geringsten OPM und damit die beste Leistungsfähigkeit in Bezug auf die Häufigkeit von Messabweichungen auf. Mit einem OPM von 5 treten hier bei 5 von 1000 Messungen bei einer Konzentration von 2 mmol/l Messabweichungen von mehr als $\pm 0,1 \text{ mmol/l}$ auf. Vergleicht man die Ergebnisse mit den Resultaten der Glucosestudie, ist Glucose mit einer Häufigkeit von weniger als 10 Abweichungen bei 1000 Messungen bei einer A-Zonen-Weite von 5 % mit den Ergebnissen von Calcium vergleichbar.

Zusammengefasst zeigten Calcium, TSH und Cholesterol die niedrigsten AZ95- und OPM-Werte und damit die beste analytische Leistungsfähigkeit. Alle 3 Analyte haben einen maximalen OPM von 50 und gleichzeitig einen maximalen AZ95 von 5 %. Dies wurde auch von Glucose in der vorangegangenen Studie erreicht (14).

Vergleichbare geringe Fehlerhäufigkeiten eines Calciumassays wurden in einer Studie von Deetz et al. veröffentlicht, bei der eine Fehlerhäufigkeit von 0,2 % mit Hilfe der Fehlerdefinition des College of American Pathologists/Clinical Laboratory Improvement Amendments (CAP/CLIA) bei 3000 Doppelmessungen eruiert wurde (8). Bei anderen in dieser Studie untersuchten Analyten war die Anzahl der Doppelmessungen teilweise sehr gering, wodurch die Aussagekraft für selten auftretende Messabweichungen eingeschränkt ist. Demgegenüber deckt Onyenekwu et al. in einer Studie mit Doppelmessungen für Calcium eine deutlich höhere Messausreißerquote von 4,9 % – ebenfalls mit der Fehlerdefinition von CAP/CLIA – auf (7). Allerdings wurde auch hier nur eine geringe Anzahl von Doppelmessungen durchgeführt. Witte et al. nimmt für die Definition einer Messabweichung ein Vielfaches der Standardabweichung und beschreibt in diesem Zusammenhang für mehrere Analyten insgesamt 447 Messabweichungen bei 100.000 Messungen bzw. 41 von 1.000.000 Messabweichungen, die potentiell den Patienten gefährden könnten (29).

Diese Beispiele verdeutlichen, wie durch Anwendung unterschiedlicher Fehlerdefinitionen die Vergleichbarkeit von Studienergebnissen erschwert wird. Das in dieser Studie angewendete Modell der A-Zone bietet daher die Möglichkeit, zum einen durch die fixen Parameter AZ95 und OPM Ergebnisse zwischen Laboratorien

und zwischen verschiedenen Analyten oder verschiedenen Assays eines Analyten zu vergleichen, gewährleistet aber gleichzeitig mittels Variierung der A-Zonen-Weite die individuelle Beurteilung eines Analyten, um hierdurch die klinischen Anforderungen und das technisch Realisierbare zu berücksichtigen. Die gewählte grafische Darstellung kumuliert alle A-Zonen-Variationen in eine Analyt-spezifische Kurve und erlaubt den Vergleich mit anderen Assays.

Zukünftige Studien sind hier erforderlich, um Daten für weitere Analyte beziehungsweise Geräte zu ermitteln. In der vorliegenden Studie wurden aufgrund der durch Doppelmessungen entstehenden Kosten und der in diesem Zusammenhang begrenzten Mittel die Untersuchungen auf neun Analyte limitiert. Für zwei Analyte wurden jeweils weniger als 10 000 Doppelmessungen durchgeführt. Die Möglichkeit seltene Messfehler aufzudecken sinkt mit abnehmender Anzahl. Durch die bei diesen Analyten ermittelte Anzahl von Messabweichungen mit einem OPM von unter 150 für ALAT und Laktat, ist die in dieser Studie durchgeführte Anzahl von Doppelmessungen (5620 bzw. 1596) allerdings ausreichend, um valide Aussagen treffen zu können.

Die hier vorgenommene Auswertung und Darstellung von Messabweichungen mit Hilfe der variierten A-Zone und den daraus abgeleiteten Kurven sowie der AZ95 und der OPM stellen eine sinnvolle Ergänzung zu Präzision und Richtigkeit bei der Bewertung der analytischen Qualität eines Assays dar (34) und vervollständigt die Aussage über dessen Leistungsfähigkeit. Insbesondere bei der Auswahl eines Messverfahrens unter Berücksichtigung der klinischen Anforderungen können die hier dargestellten Informationen über die analytische Leistungsfähigkeit eines Messverfahrens hilfreich sein.

5. Zusammenfassung

In der vorliegenden Arbeit wurden mittels Doppelmessungen Messabweichungen beschrieben, die trotz gültiger Qualitätskontrollen auftreten und deren Häufigkeit bisher nicht transparent darstellbar war. Des Weiteren wurden neue Qualitätsmarker auf der Basis von Doppelmessungen etabliert, die in Ergänzung zur Qualitätssicherung zusätzliche Informationen über die analytische Leistungsfähigkeit von Messverfahren liefern.

Die hier vorgestellten Qualitätsmarker AZ95, d.h. die Weite der A-Zone, bei der 95 % aller Doppelmessungen innerhalb der A-Zone liegen und der OPM, d.h. die Anzahl von Messfehlern pro 1000 Doppelmessungen bei einer festgelegten A-Zonen-Weite von 5 %, dienen dem besseren Vergleich der analytischen Leistungsfähigkeit von Messverfahren.

Die Kombination der Qualitätsmarker einer AZ95 von maximal 5 % und einem OPM von maximal 50 wurde in der vorliegenden Studie für drei der neun untersuchten Analyte, namentlich Calcium, TSH und Cholesterol erfüllt.

Darüber hinaus bietet das Modell der A-Zonen durch die variable Anpassung der A-Zonen-Weite die Möglichkeit, individuelle Gütekriterien für einzelne Analyten festzulegen.

Die hier neu eingeführten Qualitätsmarker stellen eine sinnvolle Ergänzung zur Bewertung der analytischen Leistungsfähigkeit von Messverfahren dar und ermöglichen die Berücksichtigung klinischer Anforderungen an ein Messverfahren und bieten somit auch eine Entscheidungshilfe bei der Auswahl eines neuen Messverfahrens.

6. Literaturverzeichnis

1. ISO 15189:2012(en), Medical laboratories — Requirements for quality and competence. Available at: <https://www.iso.org/obp/ui/#iso:std:iso:15189:ed-3:v2:en>. Accessed May 23rd 2016.
2. Westgard JO. "Westgard Rules" and Multirules - Westgard. Available at: <https://www.westgard.com/mltirule.htm>. Accessed May 23rd 2016.
3. Bundesärztekammer. Richtlinie der Bundesärztekammer zur Qualitätssicherung laboratoriumsmedizinischer Untersuchungen. Deutsches Ärzteblatt 2014;111(38):A1583-A1618.
4. CLSI. User Verification of Precision and Estimation of Bias; Approved Guideline—Third Edition. CLSI document EP15-A3. Clinical and Laboratory Standards Institute; Wayne, Pennsylvania (USA); 2014.
5. CLSI. Expression of Measurement Uncertainty in Laboratory Medicine; Approved Guideline. CLSI document EP 29. Clinical and Laboratory Standards Institute; Wayne, Pennsylvania (USA); 2012.
6. Kristiansen J. The Guide to expression of uncertainty in measurement approach for estimating uncertainty: an appraisal. Clin Chem 2003;49(11):1822–9.
7. Onyenekwu CP, Hudson CL, Zemlin AE, Erasmus RT. The impact of repeat-testing of common chemistry analytes at critical concentrations. Clin Chem Lab Med 2014;52(12):1739–45.
8. Deetz CO, Nolan DK, Scott MG. An examination of the usefulness of repeat testing practices in a large hospital clinical chemistry laboratory. Am J Clin Pathol 2012;137(1):20–5.
9. Pretorius CJ, Dimeski G, O'Rourke PK, Marquart L, Tyack SA, Wilgen U, Ungerer JP. Outliers as a cause of false cardiac troponin results: investigating the robustness of 4 contemporary assays. Clin Chem 2011;57(5):710–8.
10. Ungerer JP, Pretorius CJ, Dimeski G, O'Rourke PK, Tyack SA. Falsely elevated troponin I results due to outliers indicate a lack of analytical robustness. Ann Clin Biochem 2010;47(Pt 3):242–7.
11. Sawyer N, Blennerhassett J, Lambert R, Sheehan P, Vasikaran SD. Outliers affecting cardiac troponin I measurement: comparison of a new high sensitivity assay

with a contemporary assay on the Abbott ARCHITECT analyser. *Ann Clin Biochem* 2014;51(Pt 4):476–84.

12. Ryan JB, Southby SJ, Stuart L, Mackay R, Florkowski CM, George PM. Comparison of cardiac TnI outliers using a contemporary and a high-sensitivity assay on the Abbott Architect platform. *Ann Clin Biochem* 2014;51(Pt 4):507–11.

13. Raggatt PR. Duplicates or singletons? An analysis of the need for replication in immunoassay and a computer program to calculate the distribution of outliers, error rate and the precision profile from assay duplicates. *Ann Clin Biochem* 1989;26(Pt 1):26–37.

14. Petersmann A, Wasner C, Nauck M, Kallner A. Frequency of Extreme Differences and Clinical Performance of Glucose Concentration Measurements Judged from 21 000 Duplicate Measurements. *Clin Chem* 2013;59(6):998–1000.

15. Carraro P, Plebani M. Errors in a stat laboratory: types and frequencies 10 years later. *Clin Chem* 2007;53(7):1338–42.

16. Loh TP, Lee LC, Sethi SK, Deepak DS. Clinical consequences of erroneous laboratory results that went unnoticed for 10 days. *J Clin Pathol* 2013;66(3):260–1.

17. Bashiti O. Accidental Diagnosis of Multiple Myeloma in a 44-Year-Old White Woman due to Erroneous Results via Chemical Analyzers. *Lab Med* 2016;47(1):e5-e11.

18. Rotmensch S, Cole LA. False diagnosis and needless therapy of presumed malignant disease in women with false-positive human chorionic gonadotropin concentrations. *Lancet* 2000;355(9205):712–5.

19. Cole L, Rinne KM, Shahabi S, Omrani A. False-positive hCG assay results leading to unnecessary surgery and chemotherapy and needless occurrences of diabetes and coma. *Clin Chem* 1999;45(2):313–4.

20. Mugler K, Lefkowitz JB. False-positive D-dimer result in a patient with Castleman disease. *Arch Pathol Lab Med* 2004;128(3):328–31.

21. Roller RE, Lahousen T, Lipp RW, Korninger C, Schnedl WJ. Elevated D-dimer results in a healthy patient. *Blood Coag Fibrinol* 2001;12(6):501–2.

22. Krahn J, Parry DM, Leroux M, Dalton J. High percentage of false positive cardiac troponin I results in patients with rheumatoid factor. *Clin Biochem* 1999;32(6):477–80.

23. Ismail Y, Ismail AA, Ismail AAA. Erroneous laboratory results: what clinicians need to know. *Clin Med* 2007;7(4):357–61.

24. Ismail AA, Walker PL, Barth JH, Lewandowski KC, Jones R, Burr WA. Wrong biochemistry results: two case reports and observational study in 5310 patients on potentially misleading thyroid-stimulating hormone and gonadotropin immunoassay results. *Clin Chem* 2002;48(11):2023–9.
25. Bhat V, Chavan P, Naresh C, Poladia P. The External Quality Assessment Scheme (EQAS): Experiences of a medium sized accredited laboratory. *Clin Chim Acta* 2015;446:61–3.
26. Bailey D, Bevilacqua V, Colantonio DA, Pasic MD, Perumal N, Chan MK, Adeli K. Pediatric within-day biological variation and quality specifications for 38 biochemical markers in the CALIPER cohort. *Clin Chem* 2014;60(3):518–29.
27. Krouwer JS. Critique of the Guide to the expression of uncertainty in measurement method of estimating and reporting uncertainty in diagnostic assays. *Clin Chem* 2003;49(11):1818–21.
28. Buonocore R, Avanzini P, Aloe R, Lippi G. Analytical imprecision of lactate dehydrogenase in primary serum tubes. *Annals of clinical biochemistry* 2016;53(Pt 3):405–8.
29. Witte DL, VanNess SA, Angstadt DS, Pennell BJ. Errors, mistakes, blunders, outliers, or unacceptable results: how many? *Clin Chem* 1997;43(8 Pt 1):1352–6.
30. CLSI. How to Construct and Interpret an Error Grid for Quantitative Diagnostic Assays; Approved Guideline. CLSI document EP27-A. Clinical and Laboratory Standards Institute; Wayne, Pennsylvania (USA); 2012.
31. Cox DJ, Clarke WL, Gonder-Frederick L, Pohl S, Hoover C, Snyder A, Zimbelman L, Carter WR, Bobbitt S, Pennebaker J. Accuracy of perceiving blood glucose in IDDM. *Diabetes Care* 1985;8(6):529–36.
32. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care* 1987;10(5):622–8.
33. Parkes JL, Slatin SL, Pardo S, Ginsberg BH. A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose. *Diabetes Care* 2000;23(8):1143–8.
34. Burnett RW. Accurate estimation of standard deviations for quantitative methods used in clinical chemistry. *Clin Chem* 1975;21(13):1935–8.

A1 Abbildungsverzeichnis

		Seite
Abbildung 1	A-Zonen-Modell	4
Abbildung 2	Anzahl der Messabweichungen (bei 1000 Doppelmessungen) in Relation zur A-Zonen-Weite	11
Abbildung 3	Der AZ95 in aufsteigender Reihenfolge	12
Abbildung 4	Der OPM in aufsteigender Reihenfolge	12

A2 Tabellenverzeichnis

		Seite
Tabelle 1	Erhebungszeitraum	9
Tabelle 2	Regressionsanalyse	9
Tabelle 3	Übersicht der untersuchten Analyte	10

Original papers

Analytical robustness of nine common assays: frequency of outliers and extreme differences identified by a large number of duplicate measurements

Stefanie Neubig^{*1}, Anne Grotevendt¹, Anders Kallner², Matthias Nauck¹, Astrid Petersmann¹

¹Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany

²Department of Clinical Chemistry, Karolinska University Hospital, Stockholm, Sweden

*Corresponding author: stefanie_neubig@web.de

Abstract

Introduction: Duplicate measurements can be used to describe the performance and analytical robustness of assays and to identify outliers. We performed about 235,000 duplicate measurements of nine routinely measured quantities and evaluated the observed differences between the replicates to develop new markers for analytical performance and robustness.

Materials and methods: Catalytic activity concentrations of aspartate aminotransferase (AST), alanine aminotransferase (ALT), and concentrations of calcium, cholesterol, creatinine, C-reactive protein (CRP), lactate, triglycerides and thyroid-stimulating hormone (TSH) in 237,261 patient plasma samples were measured in replicates using routine methods. The performance of duplicate measurements was evaluated in scatterplots with a variable and symmetrical zone of acceptance (A-zone) around the equal line. Two quality markers were established: 1) AZ95: the width of an A-zone at which 95 % of all duplicate measurements were within this zone; and 2) OPM (outliers per mille): the relative number of outliers if an A-zone width of 5 % was applied.

Results: The AZ95 ranges from 3.2% for calcium to 11.5% for CRP and the OPM from 5 (calcium) to 250 (creatinine). Calcium, TSH and cholesterol have an AZ95 of less than 5% and an OPM of less than 50.

Conclusions: Duplicate measurements of a large number of patient samples identify even low frequencies of extreme differences and thereof defined outliers. We suggest two additional quality markers, AZ95 and OPM, to complement description of assay performance and robustness. This approach can aid the selection process of measurement procedures in view of clinical needs.

Key words: clinical chemistry tests; humans; reproducibility of results; analytical quality; quality control

Received: July 05, 2016

Accepted: December 03, 2016

Introduction

Internal quality control schemes (IQC), e.g. Westgard or the Guidelines for Quality Assurance of Medical Laboratory Examinations of the German Medical Association (RiliBÄK), have a long tradition in clinical chemistry to monitor the reliability of measurements and have proved to provide a reasonable quality control (1-3). The Clinical & Laboratory Standards Institute (CLSI) has advised procedures to verify the trueness and precision of routine methods (4). The schemes rule that IQC samples are measured at certain intervals, defined either by time or by number of measured patient samples. Therefore, IQC may fail to detect occa-

sional dropouts, i.e. outliers among routine samples, which may cause erroneous clinical decisions in patient care (4-15). To identify outliers, duplicate measurements may be used (16-20), but are often refrained from, due to economic pressure (21,22).

Outliers are differently defined in the literature and their frequencies are reported by various approaches (1,2,4,5,16-19,23-26). Additionally, what is perceived as an outlier may vary depending on the analyte or diagnosis (2,4,16-19,24,27). By describing magnitude and number of differences between duplicate measurements, laboratories and clinicians can define the difference they consider

an outlier and consequently estimate the frequency they will experience (27). If based on a sufficient number, duplicate measurements have the potential to reveal even low frequencies of outliers (4,20). In the present study, we report a large number of duplicate measurements in patient plasma samples under routine conditions in nine commonly used assays. It is an extension of a study published in 2013, where these investigations were performed with one glucose assay (27). To broaden the findings from the glucose study, further frequently used assays were investigated and compared. Since only measurements with valid IQC shall be included, the performance observed in the study reflects the currently accepted frequency and magnitude of extreme differences. Our aim was to detect even low frequencies of outliers and their magnitude in routine assays and to derive new quality markers for assay performance that give more detailed information than the regular IQC. Our approach aims at complementing description of performance and analytical robustness of assays.

Materials and methods

The plasma samples were from an unselected population of hospitalized patients and outpatients at the University Medicine of Greifswald (Germany). During twelve consecutive months, starting in December 2011 the catalytic activity concentrations and concentrations of aspartate aminotransferase (AST), alanine aminotransferase (ALT), and concentrations of calcium, cholesterol, creatinine, C-reactive protein (CRP), lactate, triglycerides and thyroid-stimulating hormone (TSH) were routinely measured in duplicates on a 24/7 basis. Assays were run on three Dimension Vista 1500 instruments using procedures and reagents according to the manufacturer's instruction (Siemens Healthcare Diagnostics, Eschborn, Germany). The instruments were connected to Stream-LAB (Siemens Healthcare Diagnostics, Eschborn, Germany) which allocated the samples to one of the three instruments. Once allocated to a specific instrument, the duplicate measurements were automatically ordered and performed on this same

instrument by two sample aspirations but only the first obtained result was released for patient care.

Only measurements within concentration intervals according to RiliBÄK (published in RiliBÄK's Table B1, column 4: "RiliBÄK applicable concentration intervals of columns 3 and 5"), specific for each analyte, and only measurements with valid IQC according to the RiliBÄK were included in the study (3). TRU-Liquid Monitrol (lot 1AQ104) was used as IQC material for ALT, AST, calcium, cholesterol, creatinine, lactate, Triglycerides; Liquimmune (lot 9LQ105, both Thermo Fisher Scientific, Schwerte, Germany) for TSH and Protein2 (lot 1LQH01, Siemens Healthcare Diagnostics, Eschborn Germany) for CRP. Imprecision was calculated from IQC for each analyte, IQC level (high and low) and RiliBÄK monthly control cycle. Samples were anonymized prior to data collection.

Data analysis was performed using Microsoft Excel® (2010). Ethical approval of the local ethics committee was obtained.

For each assay, differences or agreements of duplicate measurements were summarized in scatter plots including regression and correlation analyses. To categorize the observed differences between the duplicates with respect to their magnitude, a zone surrounding the equal line of the scatter plots was used (Figure 1). This area is often referred to as the A-zone (28). The A-zone (dotted lines Figure 1) is located around the equal line (hatched line Figure 1) of duplicate measurements, with the first measurement on the X-axis and the second measurement on the Y-axis. The A-zone width can be modified symmetrically around the equal line as indicated by the arrows. The triangles represent a duplicate measurement within the chosen A-zone whereas the square represents a duplicate measurement outside the A-zone.

In this study its width was systematically modified and increased to the width of 14% (Figure 2). Differences outside the chosen A-zone were regarded as outliers. Thus, the definition of an outlier depends on the width of the A-zone. A narrow A-zone consequently corresponds to strict limits and would cause small differences to be regarded as outliers whereas a wide A-zone allows large differences.

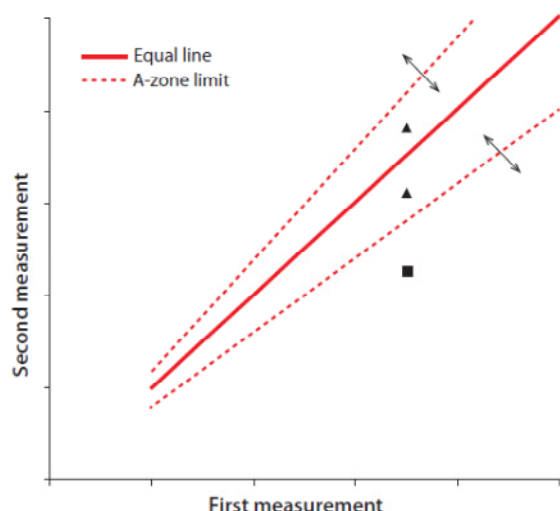


FIGURE 1. The A-zone.

The dotted lines around the equal line (hatched line) of duplicate measurements represent the A-zone, the first measurement on the X-axis and the second measurement on the Y-axis. The A-zone width can be modified symmetrically around the equal line as indicated by the arrows. The triangles represent a duplicate measurement within the chosen A-zone whereas the square represents a duplicate measurement outside the A-zone.

At each A-zone width, differences that fell outside this area were counted. The relative numbers of observations outside the various A-zones were plotted against the width of the A-zone (Figure 2).

Two quality markers can be derived from this approach and used to describe performance and analytical robustness of assays:

1. AZ95 (A-zone 95%): The width of the A-zone at which 95% of all duplicate measurements are within this zone (Figure 2, the A-zone width can be read from the x-axis where the respective analyte curve crosses the red horizontal line. This horizontal line crosses the y-axis at 50 OPM since this represents 50 out of 1000 i.e. 95%),
2. OPM (outlier *per mille*): The relative number of outliers in *per mille* if an A-zone width of 5% is used to identify outliers (Figure 2, the relative number of outliers can be read from the y-axis where the respective curve crosses the red vertical line).

The common target is 95% of the observations within an A-zone of 5%, which is already described in the CLSI EP 27 guidelines (28) and means in our study a maximum AZ95 of 5% and at the same time a maximum OPM of 50. For the comparison between different analytes it is important to consider, that individual clinical requirements induce different requirements for the AZ95 and the OPM of each analyte.

Results

The number of plasma samples run in duplicates ranged from 1596 for lactate to 73,242 for creatinine (Table 1); in total, 237,261 duplicates were measured. The imprecision calculated from the IQC from low and high levels of the IQC for each analyte is given in Table 1. At an A-zone width of 12% all assays had less than 50 measurements *per mille* outside the A-zone and showed then an as-

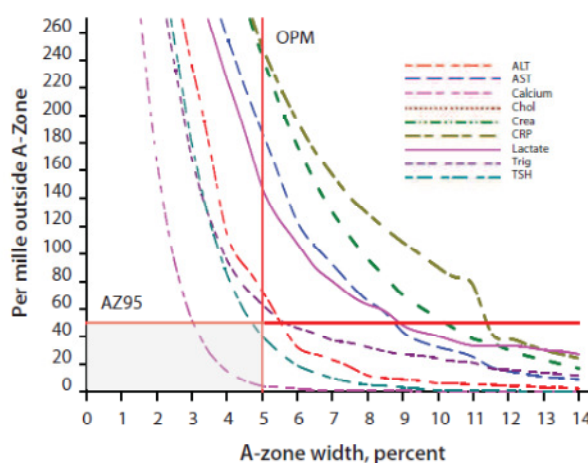


FIGURE 2. Analytical performance of nine assays. The frequency (*per mille*) of observed differences in duplicate measurements, which were outside a given width of the zone of acceptance (A-zone). AZ95, (the width of the A-zone comprising 95% of the observations) for each assay is given on the X-axis where the horizontal red line crosses the curves of the assays. The OPM (the relative number of observations in 1000 observations outside the A-zone) at an A-zone width of 5% can be read from the red vertical line. Horizontal lines crossing assay lines are in the following order: calcium, TSH, cholesterol, ALT, triglycerides, AST, lactate, creatinine and CRP. Assays that cross the shaded area in the left-hand corner comprise 95% of the observations with an A-zone width lower than 5% and have fewer than 50 *per mille* outliers at an A-zone of 5%.

TABLE 1. Basic data on studied assays and imprecision calculated from IQC.

Measurand (Measuring interval according to the manufacturer's instructions)	RiliBÄK (column 4) applicable concentration	Number of duplicates	CV (%), (target value IQC low level)	CV (%), (target value IQC high level)
ALT* (6 - 1002 U/L)	30–300	5620	3.1–5.0 (42 U/L)	2.8–3.6 (105 U/L)
AST* (3 – 1002 U/L)	19.8–400.2	11,797	2.1–5.7 (36 U/L)	0.8–1.5 (194 U/L)
Calcium* (1.25 - 3.75 mmol/L)	1.00–6.00	65,077	1.1–2.4 (2.41 mmol/L)	1.0–2.2 (2.85 mmol/L)
Cholesterol* (1.29 - 15.54 mmol/L)	1.3–9.1	12,092	1.7–5.2 (5.1 mmol/L)	1.8–5.4 (2.74 mmol/L)
Creatinine* (9 – 1768 µmol/L)	44–884	73,242	2.2–7.5 (118 µmol/L)	0.9–3.8 (629 µmol/L)
CRP† (3.1 – 190 mg/L)	1–120	27,836	2.4–6.1 (52 mg/L)	2.3–7.0 (12.1 mg/L)
Lactate* (0.1 – 15 mmol/L)	1–10	1596	1.2–4.4 (2.42 mmol/L)	1.2–8.3 (5.71 mmol/L)
Triglycerides* (0.02 - 11.3 mmol/L)	0.68–4.6	17,094	1.3–3.8 (2.35 mmol/L)	1.6–4.5 (1.35 mmol/L)
TSH‡ (0.005 – 100 mU/L)	0.1–40	23,321	2.0–4.3 (0.27 mU/L)	1.8–3.6 (16.4 mU/L)

*IQC were performed using 1AQ104 TRU-Liquid Monitor (Thermo Fisher Scientific, Schwerte, Germany).

†IQC Was performed using 9LQ105 Liquimmune (Thermo Fisher Scientific, Schwerte, Germany).

‡IQC Was performed using 1LQH01 Protein2 (Siemens Healthcare Diagnostics, Eschborn Germany).

CV – coefficient of variation. Conversion factor from SI Unit to Conventional Unit for enzymes is 60. AST - aspartate aminotransferase; ALT - alanine aminotransferase; CRP – C – reactive protein; TSH - thyroid-stimulating hormone.

ymptotical decrease. In consideration of the asymptotical curve at widen A-zone width we limited the width of the A-zone to 14%.

The number of outliers relative to the A-zone width is shown in Figure 2:

1. AZ95 is read on the X-axis in figure 2 where the horizontal red line crosses the curves of the assays. The AZ95 ranges from 3.2% for calcium to 11.3% for CRP. Triglycerides and ALT have an AZ95 of 5.5%, which is below the values found for lactate, AST, creatinine and CRP, but above those found for calcium, TSH and cholesterol.
2. OPM can be read from the vertical red line in Figure 2. An OPM of 5 *per mille* can be found for calcium and up to 250 *per mille* for CRP and creatinine. Values found for AST, lactate, creatinine and CRP indicate a poorer performance, i.e. a higher relative number of outliers than triglycerides, ALT, cholesterol, TSH and calcium.

Only three out of the investigated nine assays (calcium, TSH, and cholesterol) have

- an A-zone width equal to or lower than 5% and including 95% of the observations and
- equal to or fewer than 50 *per mille* outliers at an A-zone of 5%.

The curves of these assays cross the shaded area in the lower left corner in Figure 2.

Discussion

We used a large number of duplicate measurements to describe the performance and analytical robustness of assays.

We introduce two new quality markers for describing analytical quality (AZ95 and OPM): The AZ95 (width of the A-zone covering 95% of the observations for an assay) was chosen in analogy to the 95% confidence interval and represents the first of

the two suggested quality markers (horizontal red line, Figure 2). It ranges from about 3.2% for calcium to 11.3% for CRP.

In a clinical setting, relative terms may be difficult to handle. Therefore, we translate our findings into absolute terms. When assuming that the first measured value represents a measurement on the equal line, the AZ95 for calcium was found to be 3.2%. Therefore, at a calcium concentration of 2.0 mmol/L, 95% of all duplicate measurements could be expected between 1.94 and 2.06 mmol/L, whereas for a creatinine concentration of 100 µmol/L it would be between 90 and 110 µmol/L since its AZ95 was determined at 10.3%. Still, 5% of all measurements will deviate more. For glucose the A-zone width which comprises 95% of all observations was reported in the previous study, which used the same instrument, to be approximately 4% (27). This previously reported performance for glucose is comparable to our findings for calcium.

The second suggested quality marker OPM is the relative number of observations outside the A-zone in *per mille* at an A-zone width of 5% (vertical line, Figure 2). Calcium shows the best performance with an OPM of 5 in 1000 measurements. Translated into clinical terms: at calcium concentration of 2.0 mmol/L \pm 0.1 mmol/L the clinician would have to accept an outlier frequency of 5 *per mille*, i.e. in only 5 cases out of 1000 measurements. For CRP (e.g. at 5.0 mg/L \pm 0.25 mg/L) and creatinine (e.g. at 100 µmol/L \pm 5.0 µmol/L) the OPM is 210 and 250 in 1000 measurements, respectively (extrapolated from Figure 2). The initial study reported less than 10 *per mille* outside an A-zone width of 5% for glucose based on 21,000 duplicate measurements and therefore its performance is comparable to calcium in this study (27).

According to the quality markers, AZ95 and OPM used in the present study three assays (calcium, TSH and cholesterol) show the best analytical robustness and performance of all investigated assays with an A-zone width that comprises 95% of the observations that is lower than 5% and fewer than 50 *per mille* outliers at an A-zone of 5%. Both criteria have been reported to be fulfilled also for glucose (27). Deetz *et al.* investigated duplicate

measurements applying College of American Pathologists (CAP)/Clinical Laboratory Improvement Amendments (CLIA) error limits to identify outliers and report 0.2% outliers out of 3000 observations for calcium which corresponds to an AZ95 of about 2.5% (22). This study is in line with our findings for calcium, which showed a very low frequency of extreme differences compared to other assays. Whereas Deetz *et al.* investigated about 3000 observations for calcium, other assays had 100 observations or less (22,29). Onyenekwu *et al.* found 4.9% outliers out of 91 repeats for calcium at critical concentrations also using CAP/CLIA error limits (21). Witte *et al.* aimed to identify outliers in the sense of "errors" defined by a multiple SD e.g. 6 or 7 SD and therefore report a comparatively low frequency of 41 in one million results (0.041 *per mille*) (29).

Due to the heterogeneous approaches, results of different studies cannot be easily compared. In contrast to previous studies, our model represents a flexible approach to search for differences or outliers of various definitions by widening or reducing the A-zone accordingly. Rather than a fixed or statistically based definition of outliers, we evaluated the frequency of extreme differences of diverse magnitudes in relation to a distribution of observed duplicates around an equal line assuming identity of duplicates. In addition to precision and trueness, the frequency of outliers should be considered to describe the analytical quality of an assay (30).

Our findings describe what is presently accepted in clinical practice, i.e. the "state of the art". To facilitate comparability between laboratories and assays we suggest the fixed quality markers AZ95 and OPM in analogy to the 95% confidence interval, but our model also allows for individual adjustments. The results of our study complement performance criteria of assays and may be used for discussing potentials and limitation of assays between clinicians and laboratorians. Furthermore, this approach can aid the selection process of measurement procedures in view of clinical needs.

Due to limited resources, we focused on nine commonly used assays. The number of duplicates was below 10,000 for three assays, which limits the de-

tection of very low frequencies of extreme differences and outliers. These assays showed OPMs of approximately 55 (ALT) and 90 (lactate) which can be sufficiently identified by 5600 and 1500 duplicates, respectively.

In conclusion, duplicate measurements of large numbers of patient samples identify even low frequencies of extreme differences. We suggest two additional quality markers to describe performance and robustness of assays and report what is currently accepted in clinical practice: 1. AZ95: width of an A-zone containing 95% of all duplicate measurements, and 2. OPM: the relative number (in *per mille*) of outliers if an A-zone width of 5% is

used to identify outliers. Out of the investigated nine common assays calcium, TSH, and cholesterol have an A-zone width comprising 95% of the observations that is lower than 5% and have fewer than 50 *per mille* outliers at an A-zone of 5. Our findings complement performance criteria of assays and can aid the selection process of measurement procedures in view of clinical needs.

Acknowledgments

Reagents were partly funded by Siemens Healthcare Diagnostics.

Potential conflict of interest

None declared.

References

1. ISO 15189:2012. Medical laboratories — Requirements for quality and competence. Available at: <https://www.iso.org/obp/ui/#iso:std:iso:15189:ed-3:v2:en>. Accessed May 23rd 2016.
2. Westgard JO. "Westgard Rules" and Multirules - Westgard. Available at: <https://www.westgard.com/mltrule.htm>. Accessed May 23rd 2016.
3. Revision of the "Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations – Rili-BAEK". J Lab Med 2015;39:26–69.
4. Clinical and Laboratory Standards Institute (CLSI). User Verification of Precision and Estimation of Bias; Approved Guideline—Third Edition. CLSI document EP15-A3. Wayne, PA: CLSI; 2014.
5. Kristiansen J. The Guide to expression of uncertainty in measurement approach for estimating uncertainty: an appraisal. Clin Chem 2003;49:1822–9. <https://doi.org/10.1373/clinchem.2003.021469>.
6. Carraro P, Plebani M. Errors in a stat laboratory: types and frequencies 10 years later. Clin Chem 2007; 53:1338–42. <https://doi.org/10.1373/clinchem.2007.088344>.
7. Loh TP, Lee LC, Sethi SK, Deepak DS. Clinical consequences of erroneous laboratory results that went unnoticed for 10 days. J Clin Pathol 2013;66:260–1. <https://doi.org/10.1136/jclinpath-2012-201165>.
8. Bashiti O. Accidental Diagnosis of Multiple Myeloma in a 44-Year-Old White Woman due to Erroneous Results via Chemical Analyzers. Lab Med 2016;47:e5–e11. <https://doi.org/10.1093/labmed/lmv011>.
9. Rotmensch S, Cole LA. False diagnosis and needless therapy of presumed malignant disease in women with false-positive human chorionic gonadotropin concentrations. Lancet 2000;355:712–5. [https://doi.org/10.1016/S0140-6736\(00\)01324-6](https://doi.org/10.1016/S0140-6736(00)01324-6).
10. Cole L, Rinne KM, Shahabi S, Omrani A. False-positive hCG assay results leading to unnecessary surgery and chemotherapy and needless occurrences of diabetes and coma. Clin Chem 1999;45:313–4.
11. Mugler K, Lefkowitz JB. False-positive D-dimer result in a patient with Castleman disease. Arch Pathol Lab Med 2004;128:328–31.
12. Roller RE, Lahousen T, Lipp RW, Korninger C, Schnedl WJ. Elevated D-dimer results in a healthy patient. Blood Coag Fibrinol 2001;12:501–2. <https://doi.org/10.1097/00001721-200109000-00012>.
13. Krahn J, Parry DM, Leroux M, Dalton J. High percentage of false positive cardiac troponin I results in patients with rheumatoid factor. Clin Biochem 1999;32:477–80. [https://doi.org/10.1016/S0009-9120\(99\)00044-2](https://doi.org/10.1016/S0009-9120(99)00044-2).
14. Ismail Y, Ismail AA, Ismail AAA. Erroneous laboratory results: what clinicians need to know. Clin Med 2007;7:357–61. <https://doi.org/10.7861/clinmedicine.7-4-357>.
15. Ismail AA, Walker PL, Barth JH, Lewandowski KC, Jones R, Burr WA. Wrong biochemistry results: two case reports and observational study in 5310 patients on potentially misleading thyroid-stimulating hormone and gonadotropin immunoassay results. Clin Chem 2002;48:2023–9.
16. Pretorius CJ, Dimeski G, O'Rourke PK, Marquart L, Tyack SA, Wilgen U, et al. Outliers as a cause of false cardiac troponin results: investigating the robustness of 4 contemporary assays. Clin Chem 2011;57:710–8. <https://doi.org/10.1373/clinchem.2010.159830>.
17. Ungerer JP, Pretorius CJ, Dimeski G, O'Rourke PK, Tyack SA. Falsely elevated troponin I results due to outliers indicate a lack of analytical robustness. Ann Clin Biochem 2010;47:242–7. <https://doi.org/10.1258/acb.2010.010012>.

18. Sawyer N, Blennerhassett J, Lambert R, Sheehan P, Vasikaran SD. Outliers affecting cardiac troponin I measurement: comparison of a new high sensitivity assay with a contemporary assay on the Abbott ARCHITECT analyser. *Ann Clin Biochem* 2014;51:476–84. <https://doi.org/10.1177/0004563213499737>.
19. Ryan JB, Southby SJ, Stuart L, Mackay R, Florkowski CM, George PM. Comparison of cardiac TnI outliers using a contemporary and a high-sensitivity assay on the Abbott Architect platform. *Ann Clin Biochem* 2014;51:507–11. <https://doi.org/10.1177/0004563214534637>.
20. Raggatt PR. Duplicates or singletons? An analysis of the need for replication in immunoassay and a computer program to calculate the distribution of outliers, error rate and the precision profile from assay duplicates. *Ann Clin Biochem* 1989;26:26–37. <https://doi.org/10.1177/000456328902600104>.
21. Onyenekwu CP, Hudson CL, Zemlin AE, Erasmus RT. The impact of repeat-testing of common chemistry analytes at critical concentrations. *Clin Chem Lab Med* 2014;52:1739–45. <https://doi.org/10.1515/cclm-2014-0331>.
22. Deetz CO, Nolan DK, Scott MG. An examination of the usefulness of repeat testing practices in a large hospital clinical chemistry laboratory. *Am J Clin Pathol* 2012;137:20–5. <https://doi.org/10.1309/AJCPWPBF62YGEFOR>.
23. Bhat V, Chavan P, Naresh C, Poladia P. The External Quality Assessment Scheme (EQAS): Experiences of a medium sized accredited laboratory. *Clin Chim Acta* 2015;446:61–3. <https://doi.org/10.1016/j.cca.2015.04.007>.
24. Bailey D, Bevilacqua V, Colantonio DA, Pasic MD, Perumal N, Chan MK, et al. Pediatric within-day biological variation and quality specifications for 38 biochemical markers in the CALIPER cohort. *Clin Chem* 2014;60:518–29. <https://doi.org/10.1373/clinchem.2013.214312>.
25. Krouwer JS. Critique of the Guide to the expression of uncertainty in measurement method of estimating and reporting uncertainty in diagnostic assays. *Clin Chem* 2003;49:1818–21. <https://doi.org/10.1373/clinchem.2003.019505>.
26. Buonocore R, Avanzini P, Aloe R, Lippi G. Analytical imprecision of lactate dehydrogenase in primary serum tubes. *Ann Clin Biochem* 2016;53:405–8. <https://doi.org/10.1177/0004563215595644>.
27. Petersmann A, Wasner C, Nauck M, Kallner A. Frequency of Extreme Differences and Clinical Performance of Glucose Concentration Measurements Judged from 21 000 Duplicate Measurements. *Clin Chem* 2013;59:998–1000. <https://doi.org/10.1373/clinchem.2012.197996>.
28. Clinical and Laboratory Standards Institute (CLSI). How to Construct and Interpret an Error Grid for Quantitative Diagnostic Assays; Approved Guideline. CLSI document EP27-A. Wayne, PA: CLSI; 2012.
29. Witte DL, VanNess SA, Angstadt DS, Pennell BJ. Errors, mistakes, blunders, outliers, or unacceptable results: how many? *Clin Chem* 1997;43:1352–6.
30. Burnett RW. Accurate estimation of standard deviations for quantitative methods used in clinical chemistry. *Clin Chem* 1975;21:1935–8.